

AES 35th INTERNATIONAL CONFERENCE

Audio for Games



Steve Root
keynote speaker



Jim Anderson
AES president



Michael Kelly
conference chair



London, UK

February 11–13, 2009



The Audio Engineering Society held its first computer game-oriented conference in February, the AES 35th International Conference, *Audio for Games*. The Royal Academy of Engineering, overlooking the Royal Parks in central London, was a fantastic venue for this conference, with its history providing a contrasting backdrop to the cutting-edge technology involved in the conference topic. As is typical for England in February, the weather was cold, but conference chair Michael Kelly and the committee—Josh Reiss, Damian Murphy, Russell Mason, Pete Harrison, Rebecca Stewart, Mark Anthony, John Broomhall, Nick Laviers, Simon Goodwin, Steve Martz, and Kazutaka Someya—put together a very warm welcome for the conference attendees.

KEYNOTE SPEECH

The keynote speech of the conference was given by Steve Root of Codemasters. His main theme was that creating game audio involves the work of a wide range of people from disparate backgrounds. He highlighted the disciplines that have to work together, including musicians, academics, students, industry research and development, middleware vendors, audio programmers, and audio designers. He considered it rare for people in all these areas to meet, so he encouraged the conference attendees to make the most of this unique opportunity to share ideas between the wide range of disciplines represented.

Steve gave a list of what he thinks are the main challenges in game audio technology. Among these is the problem of providing audio in multiplayer games where the players are in the same room, in terms of how to provide individual sound scenes for each based on their individual situations within the game. Another challenge is to create full 3-dimensional audio, including how to standardize this to allow compatibility between games and systems. Yet another challenge concerns vocal morphing and filtering, both to change the characteristics of speech and to analyse and filter it appropriately (for example, a profanity filter for children). Steve finished by giving some outlandish ideas for new games that would feature audio prominently, such as music games that rate performance or compositional excellence, or games where the player acts as a hospital radio DJ to heal people, or games where the player builds architecture to transmit sound around an environment, or trace a sound to its source. Any of these (and more) would help push the profile of audio within the computer game industry.

TUTORIAL DAY

The first day of the conference was devoted to tutorials, which set the background for the more detailed events later in the conference. A wide range of topics was covered, and the differences between audio for games and audio for linear media (such as television and film) were highlighted.

Dan Bardino kicked off the day by giving a broad introduction to sound design for games, focusing on non- ➡

Authors



Among the authors presenting papers or tutorials were, from left, top row, Scott Selfon, Adam Levenson, Dan Bardino, and Leonard Paul; bottom row, Karen Collins, Jean-Frederic Vachon, Oscar Mayor, and Richard Furse.

music audio. He explained the main differences between the preplanned action of linear media and the interactive user-controlled action of games and the difficulties that these cause for the sound designer. One of the major issues is that the sound needs to react to the user actions in an appropriate way, and in a manner that is perceived to be natural. Dan used the example of a gunshot in a forest where the wildlife should react accordingly, then slowly return to their usual “background state” of activity. In addition, there are constraints imposed by the technology, such as memory and processing capacity, which the sound designer has to work around. These limitations can mean that the sound designer is restricted to using a small number of sounds, so methods are required to reduce repetition of sounds. Possible methods include separating a sound event (say, a gunshot) into small segments (such as the click of the trigger, the initial impulse of the shot, and the rumble of the reverberation), and independently manipulating each.

Adam Levenson continued by introducing the issues in producing music for games. He explained that music in games may be pre-existing and licensed for use, or it may be created specifically for the game. Starting with examples of the early bleeps of Pac-Man, he showed the evolution of music in games up to modern game music, which is highly influenced by film scores. However, as with sound design, game music needs to be able to react to the action. Early games simply switched the music depending on the action.

Subsequently an approach was used where a number of individual tracks were created that could be reproduced together and faded in or out depending on the action. Modern techniques are often more complex, with music made up of a large number of segments, some of which work together, some of which can be used to blend between other components, and some of which are individual. The game then calls these segments when necessary.

Characteristics of 3-dimensional audio in games were discussed by Scott Selfon. He explained that the screen is a very limited visual window, and that the audio can provide information about the virtual world around the player. He gave an overview of the main cues that can be manipulated to simulate aspects of sounds in space. This starts by considering position in azimuth, elevation, and distance, and then covers other aspects of sounds in virtual spaces such as the reverberation of various acoustical environments and occlusion and obstruction of the direct sound by other objects (such as walls). He played examples demonstrating the commonly available processing algorithms and showed how they can be applied to simulate audio space. He finished by mentioning a number of advanced topics that need further consideration, such as how to simulate sound effects that aren’t point sources and how to render sound in multiplayer games where each player may be in different virtual locations but are in a shared physical location.

To complement the prerecorded audio discussed in the



More authors: from left, top row, Masataka Nakahara, Snorre Farner, Inger Ekman, and Simeon Delikaris-Manias; bottom row, Steven Martz, Robert Hamilton, Blair Bitonti, and Stephen McCaul.

earlier sessions, Leonard Paul gave an overview of methods to synthesize audio. He played examples of music and sound effects created using a range of synthesis techniques, from classic additive synthesis to modern physical modeling and granular synthesis algorithms. In doing so he gave a detailed overview of the capabilities of each system in terms of the type of sound that can be created and the associated processing requirements. He focused on the possibilities afforded by physical modeling, although explained that this requires a large amount of computation. In order to make this practical for real-time synthesis within a game system, there are a range of simplification techniques available. Leonard played examples of some of these and showed how they may affect the resulting sound.

An innovative method of teaching aspects of game audio was demonstrated by Richard Stevens. Teaching material has been built into an Unreal game environment so that students can experience demonstrations of a range of audio processing techniques in an interactive manner. Richard took the audience through the game and used it to demonstrate methods to manipulate audio and to show the effect of various compromises that can be made when developing a game. Within this game environment, he demonstrated examples of most of the techniques discussed in the earlier presentations, including creation and manipulation of sound effects and variation of music based on the game action by using different transitions between musical extracts. He fin-

ished by showing a number of simple games that had been created using these techniques. Following the presentation, attendees were able to try the system for themselves.

The tutorial day was brought to a close by Jason Page who gave a Dummies Guide to Digital Signal Processing for Games. He explained that the processing power currently available in most game consoles means that audio manipulation can now be undertaken in real time, providing opportunities that were impossible only a few years ago. The main advantage of using real-time processing compared to replaying preprocessed audio is flexibility. A single sound can be processed in a wide range of ways to adapt it to different situations or different virtual environments and to reduce the potential annoyance from repetition of sounds without increasing memory requirements. However, there are difficulties with using real-time processing, including the issues of developing games for multiple platforms that have different capabilities and the competition with other aspects of the game for a share of the processing.

PAPER SESSIONS

Damian Murphy, papers chair, put together a fascinating program of presentations, made up of 23 papers and 9 posters. These covered all areas of audio for games, including sessions on spatial audio, synthesis, speech processing, and audio codecs.

The first paper session concentrated on audio effects ➡



Registration desk: Heather Lane and Pete Harrison

Conference committee: from left, Pete Harrison, events officer; Josh Reiss, conference general secretary; Michael Kelly, conference chair; Damian Murphy, papers chair; Rebecca Stewart, facilities chair; and Russell Mason, treasurer and webmaster. Simon Goodwin and John Broomhall missed this photo.

and Foley, with a focus on their ability to affect emotion. Karen Collins presented work on quantifying the affective attributes of sound effects, which takes methods developed for music semiotics and applies them to sound effects for games. Two games were created that asked the user to tag audio examples either by matching them to a picture or by adding words to describe the emotion. Based on this research, it should be possible to develop guidelines that can assist development of future game sound effects to better evoke the intended perception. Inger Ekman continued the theme by presenting work that examined how to fine-tune the emotional associations of sound effects. She discussed the relationship between the emotions arising from the story or action in the game and those arising from various aspects of perception. She summarized an experiment that looked at the effect of localization on the scariness of sound, and explained that this had a small but significant effect on the perceived emotion. The issue of repetition in games was examined by Jean-Frederic Vachon. He outlined ways to measure repetition, such as the number of times a given sound event is used and the time between each repetition. He used examples in games to highlight the problems of randomized replay of a number of different dialogue samples and methods to avoid these, such as having sound events arranged in sets where the selected set is determined by the story line of the game and the event is selected randomly from this.

The audio codecs session started with spatial coding for games. Leonid Terentiev described the Spatial Audio Object Coding system, which is based on the MPEG surround sound standard, and explained how this can be exploited in game audio. The issues of coding audio for transmitting voice communication in real time over the Internet were discussed by David Mann. He demonstrated a system that enables this for single or multichannel audio. In the third audio codecs paper, Masataka Nakahara considered the problems of level differences between games or even within a game. By measuring the audio levels in a wide range of games, he revealed the wide range of resulting levels and proposed a standard to increase consistency.

There were a number of papers related to the processing of speech and singing. Yuri Yoshida presented work on a method to manipulate a singer's identity and singing style. Related to this, Masanori Morise examined methods to extract the fundamental frequency of the singing voice. He found that by measuring the period of the vocal fold vibrations a more reliable estimation could be achieved. Oscar Mayor presented further methods to transform voice signals, which could make interesting changes such as altering gender from male to female or transforming a teenager to an older woman. And Snorre Farner demonstrated an alternative system that can be used to alter the gender and age of the voice and manipulate other qualities such as eagerness or softness.



Simon Whetham's demo included a premier of his work *Amazonas 7.1*.

The session on spatial audio covered a range of coding and spatialization techniques that can be used in audio for games. Richard Furse gave an overview of OpenAL tools, which incorporate Ambisonic coding. The main advantage of this technology is the ability to manipulate the audio in real time and render it on whichever loudspeaker or headphone arrangement the player has available. Options for loudspeaker arrangements beyond 5.1 were discussed by Simon Goodwin. He proposed an octahedral arrangement of loudspeakers intended to allow full 3-dimensional rendering of audio, but with loudspeaker positions similar to conventional 5.1 surround sound for backwards compatibility. As an alternative, Simeon Delikaris-Manias considered binaural reproduction over loudspeakers: processing audio reproduced over two loudspeakers so that the correct binaural cues are simulated at the ears of the listener. He outlined the problems that are caused by attempting this in different rooms and the methods that might be used to get around these problems. The final

spatial audio paper, presented by Dan Barry, outlined a method of upmixing material from 2-channel stereo to 5.1 surround sound. For this, the interchannel level difference of the 2-channel original was analyzed, and the results were used to remap the audio onto a surround-sound reproduction.

Robert Hamilton presented a paper on the use of the open sound control messaging protocol to create interactive musical gaming environments. He discussed how this could be used to create multiuser performance spaces with either user-centric or space-centric perspectives.

There were a number of papers that covered methods of creating and processing reverberation. Nicolas Tsingos explained the problems of prerecording reverberation (lack of flexibility) and calculating reverberation in real time (lack of processing power). He described a compromise where the early reflections and directional decay profiles are calculated in advance from a geometrical model of each virtual space, and then the dry sources are processed in real time within the game using a frequency-domain scalable processing approach. Rebecca Stewart offered another way to optimize the processing of reverberation, by analysing the impulse responses of reverberation in a wide range of positions within a space and synthesizing a generalized reverberation tail from this. Another reverberation manipulation technique was presented by Gavin Kearney. He described a method to interpolate between impulse responses captured at microphone positions in a room, which allows for greater flexibility when rendering the results. A technique to generate virtual acoustic environments was presented by Christian Borß. This takes parameters of room geometry, reverberation time, and echo density profile to create reverberation either for reproduction over headphones or over a range of loudspeaker configurations.

The final paper session of the conference concerned real-time synthesis. Niels Böttcher presented an overview of a range of sound synthesis systems that were evaluated within a game environment. He found that granular synthesis was the most successful of the methods tested, based on a combination of sound quality, realism, interaction, and preference. Kees Went gave an overview of a framework that enables research and rapid prototyping of nonlinear sound for games. It has been designed for academic researchers to adapt sound and music design in real time during active game play. A method of synthesis based on granular synthesis was demonstrated by Cécile Picard. This involves prior analysis of the audio within a game to extract suitable short segments (grains) of sound, and then using these to either recreate the original sounds or synthesize new ones according to the parameters of the game or input from the game's physics engine.

The conference also included a number of poster presentations. Simon Goodwin showed his results from investigating the reproduction systems of game consumers. He found that approximately one third used 2-channel stereo, a third used headphones, and a third used surround sound, but that 70% considered surround sound to be important to games. The majority of the posters covered spatial audio, with presenters discussing 3-dimensional rendering for mobile devices, synthesis of spatial cues, spatial sound for location-based games, and panning techniques.

SPECIAL EVENTS

The conference hosted a panel discussion on education and academic research for game audio. Panelists Mark Sarisky, Sarah LeMarie, Karen Collins, Richard Stevens, and Dave Raybould represented both academia and industry, and the discussion covered a wide range of issues including research collaboration, internships, required skills and curriculum design, and the role of accreditation.

A second panel session discussed the future of audio for games. Chaired by Adele Cutting, the panel of John Broomhall, Ciaran Rooney, Alastair MacGregor, James Slavin, and Chris Sweetman considered requirements for game audio and its associated technology. They compared the sound design of games and movies, discussing the difference between attempting natural rendition of sounds or artificial enhancement. They also explained that the mixing tools within the game are often unintuitive, and they considered the possibilities afforded by automixing. The problems of repetition of sounds in a game were examined, with the conclusion that repetition of speech was more problematic than repetition of effects. Related to this, one of the key wishes that was expressed was the need for convincing speech synthesis, which would also allow control of prosody. They agreed that current systems are a long way from sounding natural, meaning that speech needs to be recorded, bringing the associated problems of large memory requirements and repetition.

Blair Bitonti, Adam Levenson, and Stephen McCaul gave a detailed presentation on the development of the sound for Activision's *Call of Duty: World at War*. This documented all aspects of the work undertaken for the music, including details of composition and recording, and the technology in the game audio, including occlusion effects and the use of surround sound.

Modal synthesis techniques were reviewed by Francois Thibault. He highlighted the benefits, such as the reduction of memory requirements and an increase of interactivity available by controlling the parameters using the game physics engine. The session covered the main components of modal synthesis, including signal analysis and a range of real-time modal synthesis techniques.

Steve Martz and Dirk Noy gave an introduction to 7.1



Rebecca Stewart coordinated the facilities for the conference with the help of student volunteers.



Poster sessions allowed expanded interaction between authors and attendees.



From left, Roger Furness, AES executive director, Jim Anderson, president, and Karlheinz Brandenburg, governor, discuss Society activities during a coffee break.

surround sound for games. They started by giving an overview of the loudspeaker arrangement and bass management, then considered various aspects of acoustic design of replay rooms to optimize the sound.

Japanese game studios were featured in a workshop on techniques for building virtual worlds. A panel consisting of Steven Martz, Bike Suzuki, Tetsukazu Nakanishi, Kanako Kakino, Ryuichi Takada, Kazuya Takimoto, Eiji Nakamura, Chiharu Minekawa, and Kazutaka Someya (see photo below) gave an overview of the facilities employed in the creation of their games and explained the process used to develop the sound design. They showed examples of the development of ambiences and the manipulation of sound effects based on various parameters in the game, such as speed and power level. They also demonstrated how tension can be built through the menu system of a game, by introducing additional instruments to the music as the player progresses through the levels of the menu screens.

The conference also included a social evening held at

NAMCO Station, an entertainment complex with arcade games, 10-pin bowling, pool tables, and food and drink. Organized by Pete Harrison, the evening offered the participants an opportunity to let their hair down and socialize in an informal setting.

DEMONSTRATIONS

The conference also included the opportunity to experience a number of demonstrations of cutting-edge audio technology. Richard Furse demonstrated his Ambisonic-based OpenAL system, including automated tools for making acoustical measurements to inform the software of the loudspeakers in use and for rendering the audio of various games over a 14.1 surround sound system. Simon Goodwin showed the capabilities of his octahedral loudspeaker arrangement for reproducing 3-dimensional sound using six loudspeakers.

Creative use of game audio technology was demonstrated for music composition by Simon Whetham. He premiered his work *Amazonas 7.1*, which makes use of hardware and software developed for game audio as an audio-only compositional tool. A coding scheme for spatial audio was demonstrated by Ville Pulkki. His directional audio coding system can be used to render a mono recording to any multichannel reproduction system, and it includes methods to alter the spatial extent of a sound source and to generate surround sound reverb using only 2 channels.

THE FUTURE OF AUDIO FOR GAMES

The AES 35th International Conference was a fantastic event, and many attendees expressed interest in holding another conference on the topic of audio for games in the near future, to keep up-to-date with the latest advances in this exciting and rapidly developing segment of the audio industry.

Editor's note: The CD-ROM of conference papers can be purchased at <www.aes.org/publications/conf.cfm>.



The Japanese perspective on audio for games was presented in one of the conference panel discussions. Participants were, from left, front row, Kazutaka Someya and Bike H. Suzuki; back row, Chiharu Minekawa, Eiji Nakamura, Kanako Kakino, Ryuichi Takada, Tetsukazu Nakanishi, Kazuya Takimoto, and Steve Martz.